

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 08/808,031B
Source: 1FW16
Date Processed by STIC: 5/31/06

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 05/31/2006

PATENT APPLICATION: US/08/808,031B

TIME: 10:52:53

Input Set : A:\37791358.APP

Output Set: N:\CRF4\05312006\H808031B.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:

6 (i) APPLICANT: Inouye, Sumiko

7 Hsu, Mei-Yin

8 Eagle, Susan

9 Inouye, Masayori

11 (ii) TITLE OF INVENTION: PROKARYOTIC REVERSE TRANSCRIPTASE

13 (iii) NUMBER OF SEQUENCES: 54

15 (iv) CORRESPONDENCE ADDRESS:

16 (A) ADDRESSEE: DIA PIPER RUZNICK GRAY CARY

17 (B) STREET: 1650 Market Street, Suite 4900

18 (C) CITY: Philadelphia

19 (D) STATE: PA

20 (E) COUNTRY: USA

21 (F) ZIP: 19103-7300

23 (v) COMPUTER READABLE FORM:

24 (A) MEDIUM TYPE: Floppy disk

25 (B) COMPUTER: IBM PC compatible

26 (C) OPERATING SYSTEM: PC-DOS/MS-DOS

27 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

29 (vi) CURRENT APPLICATION DATA:

C--> 30 (A) APPLICATION NUMBER: US/08/808,031B

C--> 31 (B) FILING DATE: 03-Mar-1997

32 (C) CLASSIFICATION:

34 (viii) ATTORNEY/AGENT INFORMATION:

35 (A) NAME: T. Daniel Christenbury

36 (B) REGISTRATION NUMBER: 31,750

37 (C) REFERENCE/DOCKET NUMBER: 1033-CIP3-CON-03

39 (ix) TELECOMMUNICATION INFORMATION:

40 (A) TELEPHONE: 215-656-3381

41 (B) TELEFAX: 215-656-2498

44 (2) INFORMATION FOR SEQ ID NO: 1:

46 (i) SEQUENCE CHARACTERISTICS:

47 (A) LENGTH: 2176 base pairs

48 (B) TYPE: nucleic acid

49 (C) STRANDEDNESS: double

50 (D) TOPOLOGY: linear

52 (ii) MOLECULE TYPE: cDNA

55 (ix) FEATURE:

56 (A) NAME/KEY: CDS

57 (B) LOCATION: 640..2094

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

62 TCATCCGCGC GGACACCCCC TCCTACGTGC CCCCCGACGC GGAGAGCGGC GTGGAGACGG 60

RAW SEQUENCE LISTING

DATE: 05/31/2006

PATENT APPLICATION: US/08/808,031B

TIME: 10:52:53

Input Set : A:\37791358.APP

Output Set: N:\CRF4\05312006\H808031B.raw

64	TGTACCGCGT	TTCCCTGGAT	GGTCACCTGG	TGGCGGTGGA	GTGGGGCCCG	CGCACGGGCT	120
66	CGCCGCGTCA	CCAGCGGCTC	TGGTTCGACT	CGGATGCGGA	AGCCCCCGGA	GCCTACTTCG	180
68	CGCGCCTCGA	GAAGTTGGCG	GCTGACGGCT	ACATCGACGC	GGCCTCGGCA	TTGGTCTAAA	240
70	CCCTTCAACC	ACGGCTCGGC	CGCCACGCGC	GGCCGGCAGG	ACAGGTGCGA	CGAACAGACG	300
72	ACGACGTGCG	CTTCACGCGC	GAGCAGCCGA	GAGAGGTCCG	GAGTGCATCA	GCCTGAGCGC	360
74	CTCGAGCGGC	GGAGCGGCGT	TGCGCCGCTC	CGGTTGGAAT	GCAGGACACT	CTCCGCAAGG	420
76	TAGCCTGTTC	TTGGCTCTCT	CCCTCCTAGG	CACTACGGCG	AGGGTGGGTA	GCGGAGCCAA	480
78	CGACGCCACC	GCCGTTTACC	CACCCCGGCC	GTAGTGCCTA	GGAGGGGAGA	GCCGGTGAGG	540
80	CTACCGTGCC	CCAGGTAAGA	TGGTGGTGCT	TTCCCGGCCT	CCGTCGACTG	CTCGCGCCAT	600
82	GTCCCGTCTT	CCATCGCCGC	GCCCGCCCAA	GGTGACAGAC	ATG ACC GCC AGG CTG		654
83				Met Thr Ala Arg Leu			
84				1	5		
86	GAC CCG TTC GTC CCC GCA GCT TCG CCG CAG GCC GTG CCC ACG CCC GAG						702
87	Asp Pro Phe Val Pro Ala Ala Ser Pro Gln Ala Val Pro Thr Pro Glu						
88		10		15		20	
90	CTC ACC GCT CCG TCG TCA GAC GCG GCC GCG AAG CGT GAA GCC CGC CGG						750
91	Leu Thr Ala Pro Ser Ser Asp Ala Ala Ala Lys Arg Glu Ala Arg Arg						
92		25		30		35	
94	CTC GCG CAC GAA GCG TTG CTC GTC CGC GCG AAG GCC ATC GAC GAA GCG						798
95	Leu Ala His Glu Ala Leu Leu Val Arg Ala Lys Ala Ile Asp Glu Ala						
96		40		45		50	
98	GGC GGC GCC GAC GAC TGG GTG CAG GCG CAG CTC GTC TCC AAG GGG CTC						846
99	Gly Gly Ala Asp Asp Trp Val Gln Ala Gln Leu Val Ser Lys Gly Leu						
100		55		60		65	
102	GCG GTC GAG GAC CTG GAC TTC TCC AGC GCC TCC GAG AAG GAC AAG AAG						894
103	Ala Val Glu Asp Leu Asp Phe Ser Ser Ala Ser Glu Lys Asp Lys Lys						
104	70		75		80		85
106	GCC TGG AAG GAG AAG AAG AAG GCC GAG GCC ACC GAG CGC CGC GCG CTG						942
107	Ala Trp Lys Glu Lys Lys Lys Ala Glu Ala Thr Glu Arg Arg Ala Leu						
108		90		95		100	
110	AAG CGT CAG GCG CAC GAG GCG TGG AAG GCC ACG CAC GTG GGC CAC CTG						990
111	Lys Arg Gln Ala His Glu Ala Trp Lys Ala Thr His Val Gly His Leu						
112		105		110		115	
114	GGC GCG GGC GTG CAC TGG GCG GAG GAC CGC CTG GCC GAC GCG TTC GAC						1038
115	Gly Ala Gly Val His Trp Ala Glu Asp Arg Leu Ala Asp Ala Phe Asp						
116		120		125		130	
118	GTG CCC CAC CGC GAG GAG CGC GCC CGG GCC AAC GGC CTG ACG GAG CTG						1086
119	Val Pro His Arg Glu Glu Arg Ala Arg Ala Asn Gly Leu Thr Glu Leu						
120		135		140		145	
122	GAC TCC GCG GAG GCG CTG GCC AAG GCG CTG GGG CTG AGC GTC TCC AAG						1134
123	Asp Ser Ala Glu Ala Leu Ala Lys Ala Leu Gly Leu Ser Val Ser Lys						
124	150		155		160		165
126	CTC CGC TGG TTC GCG TTC CAC CGG GAG GTC GAC ACG GCC ACG CAC TAC						1182
127	Leu Arg Trp Phe Ala Phe His Arg Glu Val Asp Thr Ala Thr His Tyr						
128		170		175		180	
130	GTG AGC TGG ACC ATT CCG AAG CGG GAC GGC AGC AAG CGC ACG ATT ACG						1230
131	Val Ser Trp Thr Ile Pro Lys Arg Asp Gly Ser Lys Arg Thr Ile Thr						
132		185		190		195	
134	TCC CCC AAG CCT GAG CTG AAG GCA GCG CAG CGC TGG GTG CTG TCC AAC						1278

RAW SEQUENCE LISTING

DATE: 05/31/2006

PATENT APPLICATION: US/08/808,031B

TIME: 10:52:53

Input Set : A:\37791358.APP

Output Set: N:\CRF4\05312006\H808031B.raw

135	Ser	Pro	Lys	Pro	Glu	Leu	Lys	Ala	Ala	Gln	Arg	Trp	Val	Leu	Ser	Asn	
136			200					205					210				
138	GTC	GTG	GAG	CGG	CTG	CCG	GTC	CAC	GGC	GCC	GCC	CAC	GGC	TTC	GTG	GCG	1326
139	Val	Val	Glu	Arg	Leu	Pro	Val	His	Gly	Ala	Ala	His	Gly	Phe	Val	Ala	
140			215					220					225				
142	GGA	CGC	TCC	ATC	CTC	ACC	AAC	GCG	CTG	GCC	CAC	CAG	GGC	GCG	GAC	GTC	1374
143	Gly	Arg	Ser	Phe	Leu	Thr	Asn	Ala	Leu	Ala	His	Gln	Gly	Ala	Asp	Val	
144	230						235					240			245		
146	GTG	GTC	AAG	GTG	GAC	CTC	AAG	GAC	TTC	TTC	CCC	TCC	GTC	ACC	TGG	CGC	1422
147	Val	Val	Lys	Val	Asp	Leu	Lys	Asp	Phe	Phe	Pro	Ser	Val	Thr	Trp	Arg	
148				250						255				260			
150	CGG	GTG	AAG	GGC	CTG	TTG	CGC	AAG	GGC	GGC	CTG	CGG	GAG	GGC	ACG	TCC	1470
151	Arg	Val	Lys	Gly	Leu	Leu	Arg	Lys	Gly	Gly	Leu	Arg	Glu	Gly	Thr	Ser	
152				265					270					275			
154	ACG	CTG	CTG	TCC	CTC	CTC	TCC	ACG	GAA	GCG	CCG	CGG	GAG	GCG	GTC	CAG	1518
155	Thr	Leu	Leu	Ser	Leu	Leu	Ser	Thr	Glu	Ala	Pro	Arg	Glu	Ala	Val	Gln	
156			280					285					290				
158	TTC	CGC	GGC	AAG	CTC	CTG	CAC	GTC	GCC	AAG	GGC	CCG	CGC	GCC	CTG	CCC	1566
159	Phe	Arg	Gly	Lys	Leu	Leu	His	Val	Ala	Lys	Gly	Pro	Arg	Ala	Leu	Pro	
160			295				300					305					
162	CAG	GGC	GCC	CCC	ACG	TCG	CCC	GGC	ATC	ACC	AAC	GCG	CTC	TGC	CTG	AAG	1614
163	Gln	Gly	Ala	Pro	Thr	Ser	Pro	Gly	Ile	Thr	Asn	Ala	Leu	Cys	Leu	Lys	
164	310					315					320				325		
166	CTC	GAC	AAG	CGG	CTG	TCC	GCC	CTC	GCG	AAG	CGG	CTG	GGC	TTC	ACC	TAC	1662
167	Leu	Asp	Lys	Arg	Leu	Ser	Ala	Leu	Ala	Lys	Arg	Leu	Gly	Phe	Thr	Tyr	
168				330						335				340			
170	ACG	CGC	TAC	GCG	GAC	GAC	CTG	ACC	TTC	TCC	TGG	ACG	AAG	GCG	AAG	CAG	1710
171	Thr	Arg	Tyr	Ala	Asp	Asp	Leu	Thr	Phe	Ser	Trp	Thr	Lys	Ala	Lys	Gln	
172				345					350					355			
174	CCC	AAG	CCG	CGG	CGG	ACG	CAG	CGT	CCC	CCC	GTC	GCG	GTC	CTC	CTG	TCT	1758
175	Pro	Lys	Pro	Arg	Arg	Thr	Gln	Arg	Pro	Pro	Val	Ala	Val	Leu	Leu	Ser	
176			360					365					370				
178	CGC	GTC	CAG	GAA	GTG	GTG	GAG	GCG	GAG	GGC	TTC	CGC	GTG	CAC	CCG	GAC	1806
179	Arg	Val	Gln	Glu	Val	Val	Glu	Ala	Glu	Gly	Phe	Arg	Val	His	Pro	Asp	
180			375				380				385						
182	AAG	ACG	CGC	GTC	GCC	CGC	AAG	GGC	ACG	CGG	CAG	CGG	GTC	ACC	GGG	CTC	1854
183	Lys	Thr	Arg	Val	Ala	Arg	Lys	Gly	Thr	Arg	Gln	Arg	Val	Thr	Gly	Leu	
184	390					395					400				405		
186	GTC	GTG	AAT	GCG	GCG	GGC	AAG	GAC	GCG	CCC	GCG	GCC	CGA	GTC	CCG	CGC	1902
187	Val	Val	Asn	Ala	Ala	Gly	Lys	Asp	Ala	Pro	Ala	Ala	Arg	Val	Pro	Arg	
188				410						415				420			
190	GAC	GTC	GTC	CGC	CAG	CTC	CGC	GCC	GCC	ATC	CAC	AAC	CGG	AAG	AAG	GGC	1950
191	Asp	Val	Val	Arg	Gln	Leu	Arg	Ala	Ala	Ile	His	Asn	Arg	Lys	Lys	Gly	
192				425					430					435			
194	AAG	CCG	GGC	CGC	GAG	GGC	GAG	TCG	CTC	GAG	CAG	CTC	AAG	GGC	ATG	GCC	1998
195	Lys	Pro	Gly	Arg	Glu	Gly	Glu	Ser	Leu	Glu	Gln	Leu	Lys	Gly	Met	Ala	
196			440					445					450				
198	GCC	TTC	ATC	CAC	ATG	ACG	GAC	CCG	GCC	AAG	GGC	CGC	GCC	TTC	CTG	GCT	2046
199	Ala	Phe	Ile	His	Met	Thr	Asp	Pro	Ala	Lys	Gly	Arg	Ala	Phe	Leu	Ala	

RAW SEQUENCE LISTING

DATE: 05/31/2006

PATENT APPLICATION: US/08/808,031B

TIME: 10:52:53

Input Set : A:\37791358.APP

Output Set: N:\CRF4\05312006\H808031B.raw

```

200      455      460      465
202 CAG CTC ACG GAG CTC GAG TCC ACG GCG AGC GCC GCT CCG CAG GCG GAG      2094
203 Gln Leu Thr Glu Leu Glu Ser Thr Ala Ser Ala Ala Pro Gln Ala Glu
204 470      475      480      485
206 TGACGCTCAG CGCGCGTCCG TCGCCGACGT GCCGCGCGCC AGCAACGCCG CATTTCAGCAA      2154
208 CTCGCTCAGC CGGCGCGGGT AC      2176
211 (2) INFORMATION FOR SEQ ID NO: 2:
213 (i) SEQUENCE CHARACTERISTICS:
214 (A) LENGTH: 485 amino acids
215 (B) TYPE: amino acid
216 (D) TOPOLOGY: linear
218 (ii) MOLECULE TYPE: protein
220 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
222 Met Thr Ala Arg Leu Asp Pro Phe Val Pro Ala Ala Ser Pro Gln Ala
223 1 5 10 15
225 Val Pro Thr Pro Glu Leu Thr Ala Pro Ser Ser Asp Ala Ala Ala Lys
226 20 25 30
228 Arg Glu Ala Arg Arg Leu Ala His Glu Ala Leu Leu Val Arg Ala Lys
229 35 40 45
231 Ala Ile Asp Glu Ala Gly Gly Ala Asp Asp Trp Val Gln Ala Gln Leu
232 50 55 60
234 Val Ser Lys Gly Leu Ala Val Glu Asp Leu Asp Phe Ser Ser Ala Ser
235 65 70 75 80
237 Glu Lys Asp Lys Lys Ala Trp Lys Glu Lys Lys Lys Ala Glu Ala Thr
238 85 90 95
240 Glu Arg Arg Ala Leu Lys Arg Gln Ala His Glu Ala Trp Lys Ala Thr
241 100 105 110
243 His Val Gly His Leu Gly Ala Gly Val His Trp Ala Glu Asp Arg Leu
244 115 120 125
246 Ala Asp Ala Phe Asp Val Pro His Arg Glu Glu Arg Ala Arg Ala Asn
247 130 135 140
249 Gly Leu Thr Glu Leu Asp Ser Ala Glu Ala Leu Ala Lys Ala Leu Gly
250 145 150 155 160
252 Leu Ser Val Ser Lys Leu Arg Trp Phe Ala Phe His Arg Glu Val Asp
253 165 170 175
255 Thr Ala Thr His Tyr Val Ser Trp Thr Ile Pro Lys Arg Asp Gly Ser
256 180 185 190
258 Lys Arg Thr Ile Thr Ser Pro Lys Pro Glu Leu Lys Ala Ala Gln Arg
259 195 200 205
261 Trp Val Leu Ser Asn Val Val Glu Arg Leu Pro Val His Gly Ala Ala
262 210 215 220
264 His Gly Phe Val Ala Gly Arg Ser Ile Leu Thr Asn Ala Leu Ala His
265 225 230 235 240
267 Gln Gly Ala Asp Val Val Val Lys Val Asp Leu Lys Asp Phe Phe Pro
268 245 250 255
270 Ser Val Thr Trp Arg Arg Val Lys Gly Leu Leu Arg Lys Gly Gly Leu
271 260 265 270
273 Arg Glu Gly Thr Ser Thr Leu Leu Ser Leu Leu Ser Thr Glu Ala Pro
274 275 280 285

```

RAW SEQUENCE LISTING

DATE: 05/31/2006

PATENT APPLICATION: US/08/808,031B

TIME: 10:52:53

Input Set : A:\37791358.APP

Output Set: N:\CRF4\05312006\H808031B.raw

```

276 Arg Glu Ala Val Gln Phe Arg Gly Lys Leu Leu His Val Ala Lys Gly
277      290                      295                      300
279 Pro Arg Ala Leu Pro Gln Gly Ala Pro Thr Ser Pro Gly Ile Thr Asn
280 305                      310                      315                      320
282 Ala Leu Cys Leu Lys Leu Asp Lys Arg Leu Ser Ala Leu Ala Lys Arg
283                      325                      330                      335
285 Leu Gly Phe Thr Tyr Thr Arg Tyr Ala Asp Asp Leu Thr Phe Ser Trp
286                      340                      345                      350
288 Thr Lys Ala Lys Gln Pro Lys Pro Arg Arg Thr Gln Arg Pro Pro Val
289                      355                      360                      365
291 Ala Val Leu Leu Ser Arg Val Gln Glu Val Val Glu Ala Glu Gly Phe
292                      370                      375                      380
294 Arg Val His Pro Asp Lys Thr Arg Val Ala Arg Lys Gly Thr Arg Gln
295 385                      390                      395                      400
297 Arg Val Thr Gly Leu Val Val Asn Ala Ala Gly Lys Asp Ala Pro Ala
298                      405                      410                      415
300 Ala Arg Val Pro Arg Asp Val Val Arg Gln Leu Arg Ala Ala Ile His
301                      420                      425                      430
303 Asn Arg Lys Lys Gly Lys Pro Gly Arg Glu Gly Glu Ser Leu Glu Gln
304                      435                      440                      445
306 Leu Lys Gly Met Ala Ala Phe Ile His Met Thr Asp Pro Ala Lys Gly
307                      450                      455                      460
309 Arg Ala Phe Leu Ala Gln Leu Thr Glu Leu Glu Ser Thr Ala Ser Ala
310 465                      470                      475                      480
312 Ala Pro Gln Ala Glu
313                      485
315 (2) INFORMATION FOR SEQ ID NO: 3:
317 (i) SEQUENCE CHARACTERISTICS:
318 (A) LENGTH: 263 amino acids
319 (B) TYPE: amino acid
320 (C) STRANDEDNESS:
321 (D) TOPOLOGY: linear
323 (ii) MOLECULE TYPE: protein
328 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
330 Val Lys Leu Lys Pro Gly Met Asp Gly Pro Lys Val Lys Gln Trp Pro
331 1 5 10 15
333 Leu Thr Glu Glu Lys Ile Lys Ala Leu Val Glu Ile Cys Thr Glu Met
334 20 25 30
336 Glu Lys Glu Gly Lys Ile Ser Lys Ile Gly Pro Glu Asn Pro Tyr Asn
337 35 40 45
339 Thr Pro Val Phe Ala Ile Lys Lys Lys Asp Ser Thr Lys Trp Arg Lys
340 50 55 60
342 Leu Val Asp Phe Arg Glu Leu Asn Lys Arg Thr Gln Asp Phe Trp Glu
343 65 70 75 80
345 Val Gln Leu Gly Ile Pro His Pro Ala Gly Leu Lys Lys Lys Lys Ser
346 85 90 95
348 Val Thr Val Leu Asp Val Gly Asp Ala Tyr Phe Ser Val Pro Leu Asp
349 100 105 110
351 Glu Asp Phe Arg Lys Tyr Thr Ala Phe Thr Ile Pro Ser Ile Asn Asn

```